Nested Sparse Bayesian Learning

Ranjitha Prasad, C. R. Murthy, B. D. Rao

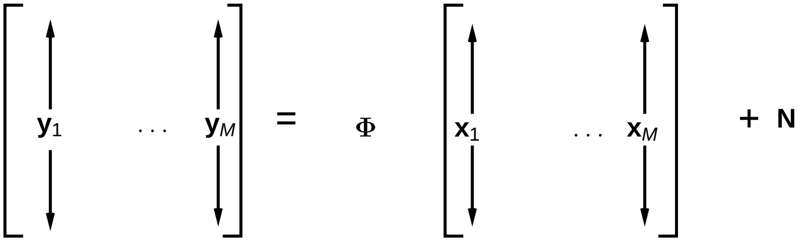
November 16, 2013

Outline

Sparse Bayesian Learning

- Problem: $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$
- Bayesian: Prior pdf on sparse vector $\mathbf{x}$
- Sparse Bayesian Learning: $\mathbf{x} \sim \mathcal{N}(0, \Gamma)$, $\Gamma = \text{diag}(\gamma)$. $|\gamma|_0 = K$
- $$\text{E : Compute } \; p(\mathbf{x}|\mathbf{y}; \gamma^{(r)}), Q(\gamma|\gamma^{(r)}) = g(\gamma, \gamma^{(r)}, \mathbf{y})$$
  $$\text{M : Compute } \; \gamma^{(r+1)} \quad (1)$$

Group-Sparsity



- $\mathbf{Y} = \Phi\mathbf{X} + \mathbf{N}$
- M-SBL solves for $\mathbf{X}$
- M-step: Solutions of the M SBL problems combined

Group-Sparsity: with and without correlation

- $\mathbf{y}_m = \Phi_m \mathbf{x}_m + \mathbf{n}$, for $1 \leq m \leq M$

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \Phi_1 & & \\ & \ddots & \\ & & \Phi_M \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} + \mathbf{n}$$

- E-step: Block (TMSBL) or Recursive (KSBL) solution
- M-step: Makes use of the structure in $\mathbf{x}$

Block Sparsity

$\mathbf{x}_{11}$	$\mathbf{x}_{12}$	$\dots$	$\mathbf{x}_{1M}$	$\dots$	$\mathbf{x}_{B1}$	$\mathbf{x}_{B2}$	$\dots$	$\mathbf{x}_{BM}$

$\mathbf{x}_1 \in \mathcal{R}^M$

$\mathbf{x}_1 \sim \mathcal{N}(0, \gamma_1 \mathbf{B})$

$\mathbf{x}_B \in \mathcal{R}^M$

$\mathbf{x}_B \sim \mathcal{N}(0, \gamma_B \mathbf{B})$

- $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$
- Algorithm: Cluster-SBL
- Block-sparsity manifested through $\gamma = [\gamma_1, \dots, \gamma_B]$

Goal of the proposed algorithms

- Breaking the problem into several sub-problems that are similar to the original problem but smaller in size
- Solve the sub-problem recursively
- Combine the subproblem solutions to create a solution to the original problem

KSBL

- Modeling correlated group sparse vectors: first order AR model
- Instead of the obtaining the MMSE solution, derive the Kalman Filter and Smoother (KFS)
- M step: Obtain γ based on KFS equations

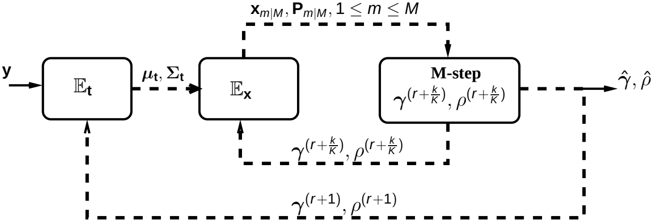
Algorithm

- Standard EM approach: Say $y_{obs} = \mathcal{M}(y_{aug})$, $\mathcal{M}(\cdot)$ is a many-to-one mapping. Standard EM:
  $Q(\theta|\theta^{(r)}) = \mathbb{E}[\ell(\theta|y_{aug})|y_{obs}, \theta^{(r)}]$,
  $\theta^{(r+1)} = \arg\max_\theta Q(\theta|\theta^{(r)})$
- In the nested approach: Say $y_{aug} = [y_{obs}, y_{mis1}, y_{mis2}]$, $y_{obs} = \mathcal{M}_1(y_{aug1})$ and $y_{aug1} = \mathcal{M}_2(y_{aug2})$. Define
  $Q_1(\theta|\theta_0) = \mathbb{E}[\ell(\theta|y_{aug1})|y_{obs}, \theta_0]$,
  $Q_2(\theta|\theta_0) = \mathbb{E}[\ell(\theta|y_{aug2})|y_{obs}, \theta_0]$ and
  $Q_{21}(\theta|\theta_{01}, \theta_{02}) = \mathbb{E}[\mathbb{E}[\ell(\theta|y_{aug2})|y_{aug1}, \theta_{01}]|y_{obs}, \theta_{02}]$
- $Q_{21}(\theta|\theta_0, \theta_0) = Q_2(\theta|\theta_0)$

Nested EM algorithm

$$\text{E} : Q_{21}(\theta|\theta^{(t+\frac{k-1}{K})}) = \mathbb{E}[\mathbb{E}[\ell(\theta|y_{aug2})|y_{aug1}, \theta^{(t+\frac{k-1}{K})}]|y_{obs}, \theta^{(t)}]$$
$$\text{M} : \theta^{(t+\frac{k}{K})} = \arg\max_{\theta} Q_{21}(\theta|\theta^{(t+\frac{k-1}{K})}) \tag{2}$$

In case of the SBL approach, $\theta = \gamma$.



Advantages

- Reduced computational complexity
- Closed form expression in the M-step
- Faster convergence

Convergence of the Nested EM approach

Theorem

Suppose $\{\theta^{(t)}, t \geq 0\}$ is a sequence in the parameter space computed with the nested EM algorithm, then $\ell(\theta^{(t+1)}) \geq \ell(\theta^{(t)})$ for each $t \geq 0$.

Proof: Sufficient to show that $Q_1(\theta^{(t+1)})|\theta^{(t)}) \geq Q_1(\theta^{(t)})|\theta^{(t)})$

Critical points

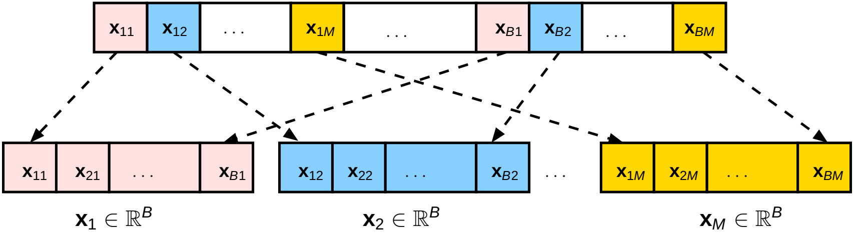
Theorem

Assuming that $Q_{21}(\cdot)$ *is continuous in its arguments, all limit points of the nested EM sequence* $\{\theta^{(t)}, t \geq 0\}$ *are critical points of* $\ell(\theta|y_{obs})$*.*

Rate of Convergence: I_{out} Vs I_{in}

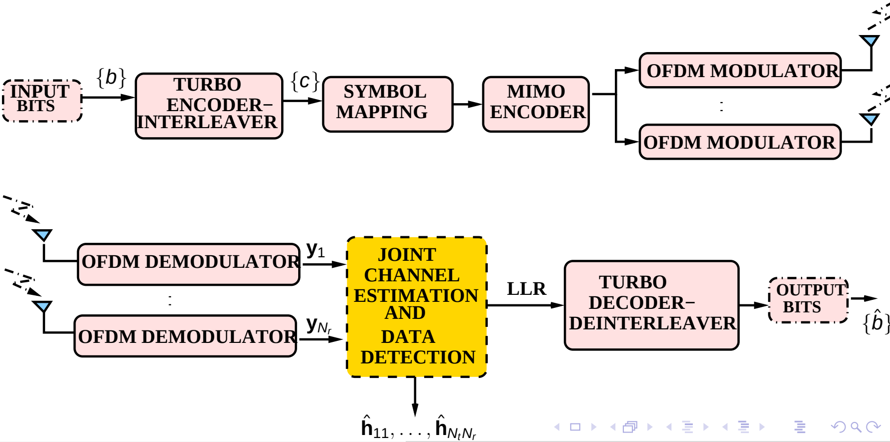
- Global rate of convergence of the nested EM approach improves with I_{in}
- In choosing I_{in}, goal is not to reach convergence, but to make progress towards the local mode
- I_{in} can vary between iterations
- If inner EM is slow to converge, large value of I_{in} is better

Application 1: Block-sparse vector recovery



- $\mathbf{y} = \sum_{i=1}^{M} \mathbf{t}_m$, $\mathbf{t}_m = \Phi_m \mathbf{x}_m + \mathbf{n}_m$
- $y_{aug1} = (\mathbf{t}, \mathbf{y})$, $y_{aug2} = (\mathbf{x}, \mathbf{t}, \mathbf{y})$

Application 2: Joint MIMO-OFDM channel estimation and data detection



- MMV system model:

$$\underbrace{[\mathbf{y}_1, \ldots, \mathbf{y}_{N_r}]}_{\mathbf{Y} \in \mathbb{C}^{N \times N_r}} = \underbrace{\mathbf{X}(\mathbf{I}_{N_t} \otimes \mathbf{F})}_{\boldsymbol{\Phi} \in \mathbb{C}^{N \times LN_t}} \underbrace{\begin{pmatrix} \mathbf{h}_{11} & \ldots & \mathbf{h}_{1N_r} \\ \vdots & \vdots & \\ \mathbf{h}_{N_t 1} & \ldots & \mathbf{h}_{N_t N_r} \end{pmatrix}}_{\mathbf{H} \in \mathbb{C}^{LN_t \times N_r}} + \underbrace{[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_r}]}_{\mathbf{V} \in \mathbb{C}^{N \times N_r}}, \qquad (3)$$

- $\mathbf{y}_{n_r} = \sum_{n_t=1}^{N_t} \mathbf{t}_{n_t n_r}, \mathbf{t}_{n_t n_r} = \mathbf{X}_{n_t} \mathbf{F} \mathbf{h}_{n_t n_r} + \mathbf{v}_{n_t}$
- $y_{aug1} = (\mathbf{t}, \mathbf{y}), y_{aug2} = (\mathbf{h}_{n_r}, \mathbf{t}_{n_r}, \mathbf{y})$